BEHAVIORAL AND BRAIN SCIENCES (1993) 16, 737–791
Printed in the United States of America

Précis of *Origins of the modern mind: Three stages in the evolution of culture and cognition*

Merlin Donald
Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6
Electronic mail: *donaldm@qucdn.queensu.ca*

Abstract: This book proposes a theory of human cognitive evolution, drawing from paleontology, linguistics, anthropology, cognitive science, and especially neuropsychology. The properties of humankind's brain, culture, and cognition have coevolved in a tight iterative loop; the main event in human evolution has occurred at the cognitive level, however, mediating change at the anatomical and cultural levels. During the past two million years humans have passed through three major cognitive transitions, each of which has left the human mind with a new way of representing reality and a new form of culture. Modern humans consequently have three systems of memory representation that were not available to our closest primate relatives: mimetic skill, language, and external symbols. These three systems are supported by new types of "hard" storage devices, two of which (mimetic and linguistic) are biological, one technological. Full symbolic literacy consists of a complex of skills for interacting with the external memory system. The independence of these three uniquely human ways of representing knowledge is suggested in the way the mind breaks down after brain injury and confirmed by various other lines of evidence. Each of the three systems is based on an *inventive* capacity, and the products of those capacities – such as languages, symbols, gestures, social rituals, and images – continue to be invented and vetted in the social arena. Cognitive evolution is not yet complete: the externalization of memory has altered the actual memory architecture within which humans think. This is changing the role of biological memory and the way in which the human brain deploys its resources; it is also changing the form of modern culture.

Keywords: cognition; cultural evolution; culture; distributed representations; external memory; human evolution; knowledge; language origins; mimesis; motor skill; neuropsychology; symbols; working memory

This book (Donald 1991) was an attempt to synthesize various sources of information – neurobiological, psychological, archaeological, and anthropological, among others – about our cognitive origins, in the belief that the human mind coevolved in close interaction with both brain and culture. I should make clear from the start that I have no illusions about my ability to become expert in all of the disciplines touched on by this enterprise; accordingly, my effort should be regarded with suspicion by all; at best, it will probably prove to be no more than a guide to some of the important questions that remain to be settled. This Précis focuses on my core theory and disregards most of the background material reviewed at length in the book itself.

My central hypothesis is that there were three major cognitive transformations by which the modern human mind emerged over several million years, starting with a complex of skills presumably resembling those of the chimpanzee. These transformations left, on the one hand, three new, uniquely human systems of memory representation, and on the other, three interwoven layers of human culture, each supported by its corresponding set of representations. I agree with multilevel evolutionary theorists like Plotkin (1988), who believe that selection pressures at this stage of human evolution were ultimately expressed and tested on the sociocultural level; hence I have described the evolutionary scenario as a series of cultural adaptations, even though individual cognition was really where the main event was taking place, since it provides the link between physical and cultural evolution. [See also Plotkin & Odling-Smee: "A Multiple Level Model of Evolution and its Implications for Sociobiology" *BBS* 4(2) 1981.]

In one sense the proposed evolutionary sequence is an exercise in interpolation not unlike many other efforts to construct a credible case for the emergence of particular morphological and behavioral features in various species. But in another sense it is a structural theory that confronts the question of how many processing levels must be interposed between the nonsymbolic cognitions of animals and the fully symbolic representations of humans. Symbolic representation is the principal cognitive signature of humans and the main phenomenon whose arrival on the scene has to be accounted for in any scenario of human evolution.

The theory posits a series of radical evolutionary changes – the punctuations, as it were, in punctuated equilibrium – rather than a continuous or unitary process. I do not rule out the possibility, indeed the likelihood, of smaller graduated changes that might also have occurred during the long period of human emergence; but judging from the anatomical and cultural remains left by hominids and early humans, the most important evolutionary steps were concentrated into a few transition

 0140-525X/93 $5.00+.00

periods when the process of change was greatly accelerated, and these major transitions introduced fundamentally new capacities.

I have made certain hard choices – for instance, I have opted for a late-language model, placing language near the end of the human evolutionary story rather than much earlier, as Parker and Gibson (1979) and Bickerton (1990) have. [See also Bickerton: "The Language Bioprogram Hypothesis" *BBS* 7(2) 1984.] For another, I have opted for a lexically driven model of language evolution, rather than placing the main emphasis on phonology, as Lieberman (1984) has, or on grammar, as Bickerton (1990) has. In fact, I have portrayed our capacity for lexical invention as a single pivotal adaptation capable of evolving into an instrument of sufficient power to support all of the higher aspects of language.

Moreover, I have postulated an early motor adaptation, intermediate between ape cognition and language, that gives primacy to the unique motor and nonverbal cognitive skills of humans. In this I am in basic agreement with Kimura (1976) and Corballis (1989; 1991), who have also argued for an early motor adaptation that preceded speech. However, I differ from these two authors in that I am much less concerned with the issue of cerebral laterality, and more focused on the representational possibilities inherent in an early motor adaptation. Moreover, I do not agree with them about the close qualitative linkage between language and serial motor skill; I see the two as qualitatively different, albeit interdependent, adaptations.

I have also tried to build a theory in such a way that specific details of chronology are not crucial to its central hypothesis, which is essentially concerned with cognitive *succession*, and consequent modern cognitive *structure*. Finally, I have chosen to construct my succession hypothesis around a fairly simple unifying theme, that of evolving cognitive architecture. This is based on my belief that brains store memories in and around their functional processors rather than somewhere else, as most computers do; and therefore that radically new representational strategies signal the likelihood of a change in the underlying neuropsychological architecture (keeping in mind that such changes may be anatomically complex).

Chronology, succession and transition. Chronology is important in that it helps us establish an order of succession and determines how many major cognitive steps were taken, and roughly when. This issue was not as difficult to resolve as one might have expected, given the controversy that seems to pursue archaeological finds. There is considerable stability in the basic number of hominid species that are currently interposed between humans and Miocene apes and in their order of appearance. Moreover, there is agreement that although australopithecines undoubtedly underwent massive anatomical and cultural change in adopting erect posture, they did not leave any evidence suggesting major cognitive evolution. There appear to be only two strong candidates for a major breakpoint in hominid cognitive evolution and these coincide with the transition periods leading to the speciation of *Homo erectus* (about −1.5 M years) and archaic *Homo sapiens* (roughly −0.3 M years), respectively. Four recent books on this subject (Bickerton 1990; Corballis 1991; Donald 1991; Lieberman 1991) have all agreed on this basic point.

The relative brain size of *Homo erectus* was much larger than that of previous hominids, eventually exceeding 70% of the modern human brain. The upward linear trend of hominid brain size accelerated sharply during the transition to *Homo erectus* and was sustained until *Homo sapiens* emerged (Lieberman 1991). This rapid increase in cerebral volume was concentrated in the association cortex, hippocampus, and cerebellum. Even taken in isolation of cultural artifacts, these anatomical facts would suggest a significant cognitive change. But the cultural evidence left by *Homo erectus* strongly confirms the presence of major cognitive evolution: *Homo erectus* manufactured quite sophisticated stone tools, devised long-distance hunting strategies, and migrated out of Africa over much of the Eurasian landmass.

A second major transition period preceded the speciation of *Homo sapiens* and was marked by another large brain expansion and the descent of the larynx. As Lieberman (1984) has argued, the latter probably coincided with the emergence of spoken language as we know it, that is, with the arrival of a high-speed vocal communication system driven by a large lexicon containing thousands of entries. Our exact lineage is still not known, but modern humans appear to have reached our present form some time prior to 45,000 years ago. All modern humans have a fully developed speech capacity, as well as complex oral cultures that incorporate myth, religion, and social ritual. This would suggest that our final period of major biological change extended over the late Middle and early Upper Paleolithic periods.

My decision to postulate a third transition in human cognitive evolution takes this scenario out of the realm of purely biological evolution toward a definition of evolution that is at once broader and more purely cognitive. If the descriptive criteria for major cognitive transitions are held constant throughout human history and prehistory, it is obvious that there have been some very major changes since the Upper Paleolithic. The likelihood that the specific mechanism of such recent change is nongenetic should not distract us from making that observation and exploring it to the fullest. Recent cognitive change is evident primarily in cultural artifacts and might have been classified along many different continua; I have singled out the development of external memory as the critical issue. The third transition seems to have started in the late Upper Paleolithic with the invention of the first permanent visual symbols; and it is still under way.

Structural issues. Here the main structural questions are: What new cognitive features were introduced at each of these three stages? And how do these three developments coalesce in the modern brain and mind and express themselves in culture? In my proposal all three stages introduced new memory features into the human cognitive system. One important consequence has been greatly improved voluntary access to memory representations; in effect, humans have evolved the architecture needed to support what Graf and Schacter (1985) have called "explicit" memory retrieval.

The first transition introduced two fundamentally new cognitive features: a supramodal, motor-modeling capacity called *mimesis*, which created representations that had the critical property of *voluntary retrievability*. The second transition added two more features: a capacity for *lexical invention* and a high-speed *phonological apparatus*, the latter being a specialized mimetic subsystem. The third transition introduced *external memory* storage and retrieval and a *new working memory architecture*.

The structural arrangement of these uniquely human representational systems is hierarchical, with mimetic skill serving as a necessary but not sufficient condition for language, while language capacity is a necessary condition for the invention of external memory devices. All of these representational systems are at the high end of the system and are aspects of what is sometimes vaguely called the "central processor" (see especially Fodor 1983; see also multiple book review: *BBS* 8(1) 1985). This proposal therefore implies that the human version of the central processor has evolved through a series of major changes and is now complex and quasimodular in its internal structure.

1. The starting point: The abilities of apes

Apes are brilliant event perceivers; as I have acknowledged in my book, they have a significant capacity for social attribution, insight, and deception, and great sensitivity to the significance of environmental events. In the latter category, I include the signing systems provided by human trainers; these are best treated as complex environmental events or challenges to which apes respond with their usual perspicuity (see, e.g., Savage-Rumbaugh 1980).

I agree with Olton's (1984) suggestion that apes have episodic memory, that is, the ability to store their perceptions of specific episodes (a position that Tulving [1984] evidently agrees with). However, they have very poor episodic recall, because they cannot self-trigger their memories: that is, they have great difficulty in gaining voluntary access to the contents of their own episodic memories *independent of environmental cues*. Thus they are largely environmentally driven, or conditioned, in their behavior, and show very little independent thought that is not directly related to specific episodes. I have called their style of thought and culture "episodic."

The limits of ape intelligence seem to be especially evident on the production side of the cognitive system. Bright as their event perceptions reveal them to be, they cannot express that knowledge. This limitation stems from their inability either to actively shape and modify their own actions or to voluntarily access their own stored representations. This might be why they cannot seem to invent gestures or mimes to communicate even the simplest intention (see, for instance, Crawford, cited in Munn 1971). They can learn signs made available by human trainers but they do not invent them on their own; nor do they seem to consciously "model" their patterns of movement, in the sense of reflecting on them, experimenting with them, and pushing them to the limits, the way humans do. This seems to indicate that they are far less developed than humans in at least two areas of motor control: the construction of conscious action models, and the voluntary independent retrieval of such models.

Without easy independent access to voluntary motor memories, even simple operations like self-cued rehearsal and purposive refinement of one's own skill are impossible, because the cognitive system remains primarily reactive, designed to react to real-world situations as they occur, and not to represent or reflect on them. Thus apes are not good at improving their skills through systematic rehearsal. The contrast with human children in this regard is striking: some apes might throw projectiles in a fight, but they do not systematically practice and improve their throwing skill the way human children do. The same applies to other kinds of voluntary action: children actively and routinely rehearse and refine all kinds of action, including facial expressions, vocalizations, climbing, balancing, building things, and so on. Although apes may have the same basic repertoire of acts, they do not rehearse and refine them, at least not on their own. In fact, it takes an incredible amount of training – on the order of thousands of trials – just to establish a single reliable naming response in chimps, and even those very context-specific responses remain reactive and episodic: for example, 97% of Kanzi's signing consists of direct requests (Greenfield & Savage-Rumbaugh 1990). Until hominids were able to model their actions and their episodic event perceptions and access those representations independently of environmental stimulation, they, like all higher mammals before them, were locked into an episodic lifestyle, no matter how sophisticated their event perceptions had become.

2. First transition: Mimetic skill and autocueing

The rationale for the first transition is based on several premises: (a) the first truly human cognitive breakthrough was a revolution in motor skill – mimetic skill – which enabled hominids to use the whole body as a representational device; (b) this mimetic adaptation had two critical features: it was a *multimodal* modeling system and it had a *self-triggered* rehearsal loop (that is, it could voluntarily access and retrieve its own outputs); (c) the sociocultural implications of mimetic skill are considerable and could explain the documented achievements of *Homo erectus;* (d) in modern humans, mimetic skill in its broadest definition is dissociable from language-based skills and retains its own realm of cultural usefulness; and (e) the mimetic motor adaptation set the stage for the later evolution of language.

(a) The primacy of motor evolution. The first really major cognitive breakthrough and the appearance of the first truly human-like species seem to have occurred with *Homo erectus*. The question commonly asked is: Do we need to attribute some language capacity to this species? This is inherently a structural question, since it asks whether language – rather, symbolic thought – is primary in the human cognitive hierarchy. Placing language this early in evolution, giving it this primacy, is a vote for a symbol-based, computational model of all human thinking.

The evidence supporting the premise that *Homo erectus* was unlikely to have had language has been reviewed in several places (most extensively by Lieberman 1984; 1991; but also by Corballis 1991; and Donald 1991). No modern investigator has argued that *Homo erectus* had speech or anything like it. However, to explain achievements like toolmaking and social coordination, several authors have attributed to *Homo erectus* a limited degree of linguistic capacity, usually labeled "proto-language" (Bickerton 1990; Parker & Gibson 1979). The current form of this notion is that *Homo erectus* had the linguistic capabilities of a two-year-old child, namely, one- and two-word utterances and intentional gesturing, but no grammar (Bickerton 1990). This seems feasible, since there have been claims that apes are very close to achieving this (see, e.g., Greenfield & Savage-Rumbaugh 1990).

There are serious problems with this position, however. For one thing, it puts the cart before the horse; it leaves out a prior motor adaptation without which language could never have evolved. In reality, apes are not even close to two-year-old children in the way they use symbols, except perhaps in their perceptions of the utility of symbols. On the motor side, they cannot even match what infants achieve during the babbling phase, let alone later on when children acquire reference, because apes cannot rehearse and refine movement *on their own*, or create models of reality *on their own*. Early language theories of evolution are seeking a "quick fix" solution, a rapid leap to some form of language without attending to the more fundamental motor changes that must have preceded it. (In the process, of course, these theories also sustain the AI agenda that attributes all higher cognition to a symbolizing process; Dennett 1992; Donald 1993.)

The primacy of motor evolution is central to any credible phylogenetic account of language. Before they could invent a lexicon, hominids first had to acquire a capacity for the voluntary retrieval of stored motor memories, and this retrieval had to become independent of environmental cueing. Second, they had to acquire a capacity for actively modeling and modifying *their own movement*. Without these two features, the motor production system could not break the stranglehold of the environment. Any language system assumes the ability of the speaker both to actively construct communicative acts and to retrieve them on demand. In other words, the system must first gain a degree of control over its own outputs before it can create a lexicon or construct a grammatical framework governing the use of such a lexicon.

This is critical from the viewpoint of cognitive theory, but one might still ask, given that an early motor adaptation was a logical necessity, couldn't it have occurred much earlier, perhaps in the australopithecines, and don't we still need language in some form to account for *Homo erectus*? I argue that a revolution in nonverbal motor skill would have had immediate and very major consequences in the realms of representation and social expression. These alone, without any further evolution, can account for the kinds of skills that have been documented in the culture of *Homo erectus;* they can also account for many of the nonverbal skills of modern humans.

There is another good reason for asserting the primacy of motor evolution: language is not the only uniquely human attribute that must be explained in an account of cognitive evolution (cf. Premack 1987). A good theory of the first cognitive evolutionary steps of humans should try to account for as many human nonverbal skills as possible. This leads to the first proposal of my theory: the first major cognitive transition broke the hold of the environment on hominid motor behavior and provided hominids with a new means of representing reality. The form of the adaptation was a revolutionary, supramodal improvement in motor control called "mimetic skill."

(b) Mimetic action as a unified supramodal system. Mimetic action is basically a talent for using the whole body as a communication device, for translating event perceptions into action. Its underlying modeling principle is perceptual metaphor; thus, it might also be called action-metaphor. It is the most basic human thought-skill, and remains fundamentally independent of our truly linguistic modes of representation. Mimesis is based in a memory system that can rehearse and refine movement voluntarily and systematically, guided by a perceptual model of the body in its surrounding environment, and it can store and retrieve the products of that rehearsal. It is based on an abstract "model of models" that allows any voluntary action of the body to be stopped, replayed, and edited under conscious control. This is inherently a voluntary access route to memory, since the product of the model is an *implementable self-image*.

The principle of voluntary retrievability, which might be called "autocueing," was thus established at the top end of the motor system. Autocueing is perhaps the most critical unifying feature of mimetic skill. Only humans can recall memories at will; and the most basic form of human recall is the self-triggered rehearsal of action, the refinement of action by purposive repetition. Purposive rehearsal reveals the presence of a unified self-modeling process, and most important, the whole body becomes a potential source of conscious representation. Retrievable body-memories were thus the first true representations, and also the most basic form of reflection, since the mimetic motor act itself represents something: systematic rehearsal "refers" to the rehearsed act itself, comparing each exemplar with a sort of idealized version of itself.

The human mimetic mechanism is supramodal at the output: that is, it can employ any part of the skeletal musculature in constructing a representation. It is supramodal at the input as well, since it can also utilize input from any major sense modality or perceptual system for its modeling purposes. A mimetic reenactment of an event – say a toolmaking sequence – might use the eyes, hands, feet, posture, locomotion, voice, or any combination of these, and the event itself might have been experienced through a variety of sensory modalities. Moreover, a given event can be mimetically represented in various acted-out versions. It follows that a mimetic act is a manifestation of a highly abstract modeling process.

The existence of a unified central "controller" for body mimesis is revealed most clearly in the unique human propensity for rhythm. Humans seem unable to resist rhythm; and even very young children spontaneously imitate, rehearse, and modify the rhythmic sounds and actions around them, with varying degrees of sophistication. The transferability of rhythm to virtually any skeletomuscular system in the body reveals the abstractness of human mimetic action-modeling: rhythms can be transferred from one muscle group to another, singly or in combination. For example, a sound rhythm initially modeled by the fingers can be transferred to the feet, or to the axial locomotor systems (as it is in dance), or to the head, face, eyes, tongue, voice; or to any subset of these in combination. Rhythm is thus an excellent paradigm for mimetic skill, in which an abstract perceptual event (usually a temporal pattern of sound) is "modeled" by the motor system.

Note that this modeling process relies on a principle of resemblance by which some property abstracted from sound is reproduced in motion; but these "resemblances" can be very indirect and elaborate, and innovation and mimetic "wit" are evident in more sophisticated human rhythmic games. Thus the modeling process is metaphoric or holistic: many variants of the basic rhythm may meet the criterion of resemblance, and the rhythm itself is not easily reduced to digital or discrete units combined according to "rules"; rather, it is the Gestalt, or overall pattern, that is primary.

Human mimetic capacity extends to larger time scales; it extends to the purposive sequencing of larger chunks of body movement over much longer periods of time. This assumes an extended mimetic imagination capable of imagining a series of actions in environmental context. If hominids could visually track and "parse" a complex event, as well as apes, say, then given the location of the mimetic controller at the top end of the event-perception system they should have been able to reenact complex events once large-scale action-modeling was within the capacity of the motor system.

(c) Sociocultural implications of mimetic action. An improvement of this magnitude in primate motor skill would inevitably have resulted in changes to hominid patterns of social expression. Existing repertoires of expressions would have become raw material for this new motor-modeling mechanism. By "parachuting" a supramodal device like mimesis on top of the primate motor hierarchy, previously stereotyped emotional expressions would become rehearsable, refinable, and employable in intentional communication. This would allow a dramatic increase in the variability of facial, vocal, and whole-body expressions as well as in the range of potential interactive scenarios between pairs of individuals or within larger groups of hominids. It is important to note that since a supramodal mimetic capacity would have extended to the existing vocal repertoire, it would have increased selection pressure for the early improvement of mimetic vocalization, a skill whose modern residue in speech is known as prosody.

Given a mechanism for intentional rehearsal and refinement, constructional and instrumental skills would also have moved to another plane of complexity through sharing and cultural diffusion. Improved toolmaking was in many ways the most notable achievement of *Homo erectus*, but it is important to realize that the manufacture of a new kind of tool implies a perceived need for that tool and corresponding advances in both tool use and pedagogy. Mimetic skill would have enabled widespread diffusion of new applications as well as supporting the underlying praxic innovations that led to new applications.

In addition to toolmaking and emotional expression, motor mimesis would have allowed some degree of quasi-symbolic communication, in that it would have allowed hominids to create a very simple shared semantic environment. The "meaning" of mimed versions of perceptual events is transparent to anyone possessing the same event-perception capabilities as the actor; thus, mimetic representations can be shared and constitute a cognitive mechanism for creating unique communal sets of representations. The shared expressive and social ramifications of mimetic capacity thus follow with the same inevitability as improved constructive skill. As the whole body becomes a potential tool for expression, a variety of new possibilities enters the social arena: complex games, extended competition, pedagogy through directed imitation (with a concomitant differentiation of social roles), a subtler and more complex array of facial and vocal expressions, and public action-metaphor such as intentional group displays of aggression, solidarity, joy, fear, and sorrow. These would perhaps have constituted the first social "customs," and the basis of the first truly distinctive hominid cultures. This kind of mimetically transmitted custom still forms the background social "theater" that supports and structures group behavior in modern humans.

Greater differentiation of social roles would also have been made possible by mimetic skill. The emergence of mimetic skill would have amplified the existing range of differences between individuals (and groups) in realms such as social manipulation, fighting and physical dominance in general, toolmaking, tool use, group bonding and loyalty, pedagogical skill, mating behavior, and emotional control. This would have complicated social life, placing increased memory demands on the individual; but these communication tools would also have created a much-increased capacity for social coordination, which was probably necessary for a culture capable of moving a seasonal base camp or pursuing a long hunt.

It is important to consider the question of the durability of a hominid society equipped with mimetic skill: adaptations would not endure if they did not result in a stable survival strategy for a species over the long run. Mimesis would have provided obvious benefits, allowing hominids to expand their territory, extend their potential sources of food, and respond more effectively as a group to dangers and threats. But it may also have introduced some destabilizing elements, especially by amplifying both the opportunities for competition and the potential social rewards of competitive success in hominid culture.

(d) The dissociability of mimetic skill in modern humans. The neuropsychological dissociability of mimesis can be demonstrated from neuropsychological studies of modern humans with brain injury. Certain paroxysmal aphasics manifest a unified, coherent strategy for dealing with reality that has the properties of a purely mimetic strategy. Their cognitions have a style that is often (I believe simplistically) termed a "right-hemisphere" strategy; it shows a degree of unity and a complete independence from language that must be explained.

One well-documented case, Brother John (Lecours & Joanette 1980), suffered from seizures lasting as long as ten or eleven hours, during which all aspects of language – including inner speech – were "excised" from his mind. Nevertheless, he remained fully conscious, able to find his way around, able to operate an elevator or a radio (he used the news station on the radio to test whether he was regaining speech comprehension), and capable of communicating with gesture and mime. Most important, he retained perfect episodic recall for most of these seizures: he could remember what went on during the spell, including who entered and left the room and what he had done with his time. This implies that neither his formation of retrievable episodic memories nor his subsequent retrieval of them could have depended upon having a functional language system at the time of storage. Nor could his functioning mimetic skills have depended on language. There are other neurological syndromes that produce a somewhat similar profile – some cases of global aphasia, for instance – but most patients suffer from other disabilities as well as permanent impairments and this makes clear distinctions between language and nonlanguage symptoms difficult to derive. The uniqueness of paroxysmal cases lies in their lack of nonlanguage symptoms such as apraxia, agnosia, amnesia or dementia, and their ability to return to normal after the seizure.

Further evidence comes from documented histories of deaf-mute people raised in hearing communities without formal sign language training. Such individuals could have had none of the lexical, syntactic, or morphological features normally associated with language. They obviously lacked a sound-based lexicon of words; they couldn't read or write, and had no access to a community of other deaf individuals who signed, and thus also lacked a visually based lexicon. Yet, by some accounts (e.g., Lane 1984) such individuals retained a capacity for all aspects of what I have identified as mimetic cognition: a full range of human emotional expressivity, gesture, mime, dance, athletic and constructional skills, and an ability to participate in reciprocal mimetic games.

The persistence of mimetic skill is evident in modern society. In fact, the realm of mimetic representation is still relatively autonomous from that of language and remains essential to the training of those who work with the body, such as actors or athletes, as well as to those who practice traditional constructional skills, such as arts and crafts. It is central to human social effectiveness and to the practice and teaching of games, competitive skills, and many group expressive customs, as for instance in the intentional use of group laughter as punishment, or the signaling of deference, affection, manliness, celebration, and grief; or the maintenance of group solidarity (see, for instance Argyle 1975; Ekman et al. 1969; Eibl-Eibesfeldt 1989). [See also Eibes-Eibesfeldt: "Human Ethology" *BBS* 2(1) 1979.]

(e) Mimesis as a preadaptation for language. Mimetic skill was, fortuitously, an important preadaptation for the later evolution of language. It allowed hominid tool technology and social organization and the shared realm of custom and expression to become more complex. Given the inherent fuzziness and ambiguity of mimetic representation, it would eventually have reached a level of complexity where a method of disambiguating intended mimetic messages would have had immediate adaptive benefits. Thus it created conditions that would have favored a communication device of greater speed and power.

On a more fundamental level, however, the principle of self-triggered voluntary retrieval of representations had to be established in the brain before the highly complex motor acts of speech would have been possible. Phonetic skill has been called "articulatory gesture" by various investigators (Brownman & Goldstein 1989); the whole higher apparatus of speech depends on the basically mimetic ability of individuals to create rehearsable and retrievable vocal acts, usually in close connection with other mimetic acts. In a word, language per se is layered on top of a mimetically skilled phonological system.

Language is not confined to the phonological system, however, because mimesis is inherently supramodal; thus, when phonology malfunctions, other mimetic subsystems may be harnessed by the language system. This is particularly visible in Petitto and Marentette's (1991) elegant documentation of infant babbling in sign-language environments, which occurs at exactly the same time as phonological babbling and has the same properties. Deaf infants growing up in deaf-signing households showed themselves to be very good at miming the motor *principle* behind signing, if not the signs themselves; that is, their manual "babbling" reflected the conditional probabilities of their expressive environments on a purely mimetic level. This is exactly what babbling infants do in hearing households: they model, in their actions, one of the most obvious dimensions of motor behavior to be observed in their families: repetitive, and to the infant, apparently random production of phonological acts.

Babbling, whether oral or manual, is reference-free in the linguistic sense – that is, it has no linguistic meaning – but it is nevertheless truly representational, in that babbling patterns are (eventually) excellent motor models of the expressive patterns the infants observe around them. They reproduce not only the elementary units of language, but also the larger mimetic envelope of expression as well: for example, prosody, and the habit of alternation, or "waiting one's turn" in expressive exchanges. Since babbling is free of linguistic reference, the brain mechanism that supports it does not have to be linked to language per se; rather, these eight- to ten-month-old infants look very much like good supramodal mime artists. And the supramodal nature of their babbling is very revealing: the fact that babbling isn't confined to phonology suggests that a supramodal mimetic adaptation evolved first, with phonology developing later as a specialized subsystem of mimetic capacity.

There are other theories that view early advances in praxic skill as preadaptations for language (Corballis & Beale 1976; Kimura 1976). Kimura observed that oral and manual apraxia and aphasia often result from the same left-sided lesions; from this she inferred that language and voluntary movement control are linked, possibly to a common processor. However, the neuropsychological case for linking pantomime and language to the same left-sided serial processor has since disintegrated (see Poizner et al. 1987; also Square-Storer et al. 1990).

These authors did not provide any theoretical justification of why praxis should have been an essential preadaptation for language, but Corballis's more recent (1989; 1991) hypothesis faces this problem squarely. His idea is that the left hemisphere acquired a general-purpose capacity for "generativity" that served as the common substrate for image generation and praxis and later for language. Generativity requires categorical perception (Harnad 1987), or the decomposition of the object world into elementary units; and it also requires the ability to recombine these units, as in both phonology and image generation (Kosslyn 1988). In Corballis's view, these two aspects of generativity evolved for improved praxis, forming a preadaptation for the later emergence of language. A closely related theory has been proposed by Greenfield (1991), who argues for a common left-sided mechanism for combinatory praxis and phonology, at a prelinguistic level.

The concept of mimetic skill proposed here differs fundamentally from both Corballis's idea of generativity and Greenfield's left-sided praxic "module" in its reduced emphasis on cerebral laterality; as pointed out at some length in my book, mimetic skill is probably bilateral (which is not to say that it is symmetrical) in distribution. More important, it differs in the nature of the proposed underlying mechanism. Generative praxis is conceived of as categorical, rule-governed, and serial in its manner of operation, whereas mimesis is basically a holistic or analog system that can model over time as well as space. A capacity for serially recombining categorical units would not easily account for the complex, fuzzy, holistic process of comparing movements against their idealized versions (cf. Moerck 1989), or of producing event reenactments, as in charades or pantomime. The generative modeling of the mimetic action-patterns that humans create and refine (including phonology) seems far too metaphorical and analog in principle to fit easily into this kind of quasisymbolic computational framework.

3. Second transition: Lexical invention

The rationale for the second transition is, briefly, as follows: (a) since no linguistic environment yet existed, a move toward language would have depended primarily on developing a capacity for *lexical invention;* (b) phonological evolution was accelerated by the emergence of this general capacity for lexical invention, and included a whole complex of special neuronal and anatomical modifications for speech; (c) the language system evolved as an extension of lexical skill and gradually extended to the labeling of relationships between words and also to the imposition of more and more complex metalinguistic skills that govern the uses of words; and (d) the natural collective product of language was narrative thought (essentially, storytelling), which evolved for specific social purposes and serves basically similar purposes in modern society.

(a) Lexical invention. Lexical invention is not yet understood in terms of mechanism. There is no viable computational model of this process and neural network models have not yet reached the point where anything so complex could be simulated. The process mapping the "lemma" or meaning-based side of the lexicon onto the form of the symbol – whether it is phonological or manual – involves much more than the association of a discrete signifier, or form, with a discrete meaning. The previous section argues that phonology, like any mimetic system, works according to a metaphorical principle; but so does lexical invention, if Wittgenstein (1992) or Johnson-Laird (1983) are to be believed. In other words, both word forms and meanings tend to be fuzzy, and neither side in the lexical entry is cleanly defined or discrete. Nevertheless, the tension between word form and meaning is a creative one that greatly increases the range of things that can be represented.

As discussed at some length in the book, the invention of a symbol is a complex process that involves labeling and differentiating our perceptions and conceptions of the world, including other symbols as parts of that world. Form is mapped onto meaning, but meaning is *defined* by that same process, in a reciprocal tension. This reciprocal tension is evident even now, after at least 45,000 years of lexical invention. Languages are constantly changing their particular mappings of form onto meaning; for instance, all of the tremendously diverse aboriginal American languages derived from three root Asiatic languages within the past 15,000 years (Greenberg & Ruhlen 1992); and the entire Indo-European group of languages, including language groups as diverse as Sanskrit, Gaelic, Latin, and Greek, have all evolved from a common ancestor within the past 7,000 years (Renfrew 1989). This incessant pattern of change suggests that the driving force behind lexical invention – the need to define and redefine our maps of meaning onto word forms – is more fundamental and considerably less rigid than the specific forms and rules of language at any given moment.

(b) The phonological adaptation. Phonology was not *the* primary language adaptation, but rather a specialized mimetic subsystem that supported the primary adaptation, lexical invention. The specialized anatomical subsystem that supports phonology evolved after the evolution of a supramodal lexical capacity, or more properly perhaps, concurrently with it, in a mutually reinforcing manner. As a mimetic subsystem, phonology has the same basic properties as all mimetic action, such as rehearsability, autocueing, and purposive refinement. The fact that language can be offloaded to other motor modalities, as it is in the sign languages of the deaf, is evidence of the secondary position of phonology in the evolutionary chronology. Phonology by itself could not have created a lexicon, and without lexical invention it is doubtful whether humans would have been subjected to selection pressures favoring such a powerful phonological system.

Nevertheless, it was a very complex and important adaptation, and without it, archaic sapient humans may not have been able to sustain the expansion of lexical capacity that they subsequently did. The great survival value of phonology to archaic humans is evident in the fact that it evolved despite the great respiratory dangers associated with a descended larynx, and in the sheer anatomical complexity of the adaptation. Included in the phonological adaptation were (as a minimum): the descent of the human larynx and the redesign of the supralaryngeal vocal tract, with corresponding central motor programming devices; a specialized auditory device for achieving improved auditory object constancy, which feeds back directly onto speech production; the articulatory loop, for immediate literal recall of articulated messages; and a specialized, large-capacity auditory memory system of word forms (see Levelt 1989; Lieberman 1975; 1984).

The importance of phonology should not be underestimated. There is little alternative to the notion that the original form of language is spoken language. There is an easy relationship between vocalizing and language, perhaps because phonology is fast, highly portable, and less likely to interfere with locomotion or praxis: in this sense it is a special "channel" of communication that can float freely above the largely visuomotor world of events, constructing commentaries unimpeded. Moreover, it works at a distance and in the dark, two features that have great adaptive value.

Phonology also has the special virtue of being able to generate a virtually infinite number of easily retrievable sound patterns for symbolic use. Human retrieval capacity for oral words is extraordinary; we carry around tens of thousands, and in the case of some multilinguals, hundreds of thousands of words; most other species, from bees to the Great Apes, seem to be limited to a few dozen expressions at most in the wild; and this limitation even applies to Cheney and Seyfarth's (1990; see also multiple review: *BBS* 15(1) 1992) vervet monkeys.

The natural dominance of phonology over manual signing is evident in experimental settings where subjects are encouraged to tell stories about specific experiences and their gestures are videotaped. In such experiments (cf. McNeill 1985), facial and manual gesture fall into a secondary support role, their timing ruled to the millisecond by spoken words; even the mimetic dimension of voice, prosody, remains secondary to the expressed meaning. Phonology is thus clearly the medium of choice for language itself. It should be added, however, that this pattern of dominance is often broken, especially in humor (including the humor of children), where the semantic counterpoint between what is said and what is gestured or done can become a powerful means of expression in itself. The ease with which humans can parallel-process two contradictory messages – mimetic and linguistic – has been long known to playwrights.

It is important to note that these new representational acts – speech and mimesis – can be performed covertly as well as overtly. Covert speech has been called "inner speech" by Baddeley (1986), who considers it to be equivalent to the activation of the central aspects of articulation, without actual motor execution. The mental operation we call "imagination" can similarly be seen as mimesis without motor execution of the imagined acts. The control of mimetic imagination (probably even of visual generative imagery, which is facilitated by imagined self-movement) presumably lies in a special form of kinematic imagery. Autoretrievability is just as crucial for covert imaginative or linguistic thought as it is for the overt, or acted-out equivalent. Thus, given a lexicon, the human mind became able to self-trigger recall from memory in two ways: by means of mimetic imagination and by the use of word symbols, either of which could be overt or covert.

(c) Grammar and metalinguistic skill. I have opted for a lexically driven, rather diffuse model of language evolution partly because it fits in well with the preceding evolutionary scenario and partly because I judge this to be a scientific "best bet" on the basis of an extensive review of neurolinguistic research. The main issue here is whether grammar and metalinguistic skills such as those which operate at the level of discourse and logic require a separate adaptation in addition to phonology and lexical invention. If one were to try to envisage language in such a way as to meet all of Fodor's (1983) requirements for a true linguistic "module," a separate grammar module would surely fail on a number of counts (as Fodor has acknowledged), especially inasmuch as it appears to be completely interpenetrable with the rest of language, and closely tied to semantics.

Moreover, the neurological case for a separate grammar module is weakened by recent cross-linguistic studies of aphasia, which strongly suggest that there is no specific brain lesion, nor any specific pattern of grammatical deficit, that is universally found in agrammatics of all languages. According to the "competition" model proposed by Bates and MacWhinney (1987), the whole perisylvian region of the left hemisphere is diffusely dedicated to language, function words and grammatical rules being stored in the same tissue as other kinds and aspects of lexical entries. However, I readily admit that this issue, like many others in this field, is still not conclusively resolved; there is electrophysiological evidence that function words – those most relevant to grammar – might have a different cerebral representation from open-class words (Neville 1992).

A related issue is whether it seems necessary a priori to posit a separate adaptation for the invention and transmission of grammar and the metalinguistic skills that support extended discourse. I have cited some biographical accounts of symbolic invention in mathematics as examples of how difficult it is to invent any new word or symbol. Once a concept has been "captured" lexically for the first time, it seems to become much easier to transmit it to others, but its original invention is generally difficult; and every new invention must then be subjected to the shared linguistic market for acceptance or rejection. All original lexical invention is difficult and collective. This applies to all classes of words, and there is no compelling reason why closed-class, or function, words might not be viewed as a product of the same skill that enabled the invention of nouns and adjectives. If the lexical inventor can isolate and define an abstract concept like "run" (including the exclusion rules for its correct use), it is not clear why that same mind could not isolate and define a function word like "with." Grammatical concepts do not seem to demand special treatment in their invention and, presumably, their transmission.

It would certainly make good evolutionary sense to attribute language to one core adaptation whose further evolution could account for all the attributes of language and language-based thought. Part of the reason is, simply, time; there wasn't enough time in the human story for more radical cognitive adaptations, and a capacity for lexical invention is the obvious *sine qua non* of language, and hence must be put first. If we put lexical invention earlier in the scenario, as Bickerton (1990) did, there might have been time for a separate adaptation for grammar; but then we must explain why a powerful capacity like lexical invention would have evolved when it was apparently not needed 1.5 million years ago, and then failed to develop further for over a million years. If we put lexical invention late, as I and many others do, there doesn't seem to be enough time to allow for a second adaptation for grammar.

(d) Sociocultural ramifications. Spoken language increased the number and complexity of available words and grammars and altered human culture by introducing a new level of shared representation. My hypothesis is that mimetic skill continued (and continues) to serve its traditional social purposes perfectly well: it still provides the cognitive foundation for institutions like dance, athletics, craft, ritual, and theater. Oral language initially carved out its own sphere of influence within mimetic culture, eventually assuming a dominant and governing role in human culture, but never eliminating our basic dependence on mimetic expression. Oral language remains focused on the human world, particularly on relationships (Dunbar 1993), and this pattern extends to a wide range of cultures, from technologically primitive hunters and gatherers to highly urbanized modern European societies.

The natural product of language is narrative thought; in this sense, language, like mimesis, evolved primarily as a method of modeling reality. Dunbar (1993) has argued that the normal social use of language is storytelling about other people – gossip – and he has produced observational data to prove this. But day-to-day storytelling in a shared oral culture eventually produces collective, standardized narrative versions of reality, particularly of past events; and these become what we call the dominant "myths" of a society. It is interesting that all documented human societies, even the most technologically primitive ones, have elaborate systems of myth, which appears to reflect the earliest form of integrative thought. These socially pervasive constructs continue to exert a major influence on the way oral societies – and indeed most modern societies – are run: thus I have suggested that many cultures might be labeled "mythic," after their governing representations.

4. Third transition: The externalization of memory

The case for a third cognitive transition is based on arguments, partly structural and partly chronological, that are similar in principle to those used for the first two; but the physical factors that supported the third transition are a little different, inasmuch as the latter was driven primarily by technological rather than biological developments. The chronological evidence is based on the rapid emergence of whole new classes of memory representations – external memory records – as well as a major change in the types of symbolic artifacts produced by humans. The structural argument is based partly on neurophysiological and neuropsychological evidence bearing on localization and plasticity and partly on an analysis of cognitive architecture in the context of our new relationship with external memory.

The historical case for a third transition rests on evidence that since the Upper Paleolithic humans have gradually developed three new representational devices. The first was *visuosymbolic invention*, which advanced through various well-documented stages, culminating in a variety of complex graphic and numerical conventions and writing systems. The second was *external memory*, which evolved to the point where external memory records, mediated by a "literate" class, started to play a governing role. The third was the emergence of very large, externally nested cultural products called *theories*. I will not reiterate the voluminous historical evidence for this, partly for reasons of space, but mostly because my chronology is neither original nor in dispute. The real argument for grouping together these historical trends into a so-called third transition is a structural one.

The structural case can be stated as follows: (a) external memory has introduced radical new properties into the collective storage and retrieval systems of humans; (b) the use of these external storage systems is difficult and requires a major redeployment of cerebral resources toward establishing literacy-related "modules" in the brain; (c) the physiological basis for this reorganization probably lies in neuronal epigenesis and plasticity; and (d) the role of biological working memory has been changed by the heavy use of external memory.

(a) New properties. Early humans, like their primate predecessors, depended heavily on their natural or biological memory capacities. Even though mimetic skill and language enabled humans to create a shared representational culture, the actual physical storage of that collective knowledge depended on individual memory. Thought was dependent on biological working memory, and whatever was seen or heard had to be remembered and rehearsed either in imagination or in speech. The contents of our long-term store were accessible only by means of the limited associative strategies available to biological memory, such as similarity and contiguity; thus, the need for oral mnemonics, extensive literal oral recitation, and a dependence on specialized individuals, like shamans, to preserve particularly important memory material.

The advantages of external memory are easily documented. External symbolic storage systems allow humans to circumvent, at least partially, the limitations of biological working memory, while creating a wide range of new storage, retrieval, and processing possibilities. By changing the physical medium of storage, human memory systems have acquired new properties, especially retrieval properties. I have suggested the term "exogram" to complement the notion of a biological "engram." As shown in Table 1, exograms introduced new possibilities into the human representational universe.

Exographic storage constitutes a hardware change just as real as the biological hardware changes that mediated the first two transitions; and its effect on overall memory structure may have been even greater. The exportation of memory storage has literally meant that the human race, as a collectivity, can now evolve new memory systems at the accelerated rate of technological change, as opposed to the relatively slow rate of genetic change. Perhaps the most important new features introduced by external storage are radically different options in memory retrieval, and the fact that exograms are easily reformattable. Extensive reformatting can modify the kinds of ideas and images that are available to store in biological memory, and so on, in iterative loops. This iterative crafting of complex memory records has produced completely new types of symbolic representations that had no equivalents in preliterate oral cultures – examples might include the servicing manuals for a rocket engine, the equations proving the Pythagorean theorem, a corporate income tax handbook, a heat-map of the troposphere, or the libretto and score for *Eugene Onegin*.

Table 1. *Some properties of engrams and exograms*

Engrams	Exograms
internal memory record	external memory record
fixed physical medium	virtually unlimited media
constrained format	unconstrained and reformattable
impermanent	may be permanent
large but limited capacity	virtually unlimited
limited size of single entries	virtually unlimited
not easily refined	unlimited iterative refinement
retrieval paths constrained	retrieval paths unconstrained
limited perceptual access in audition, virtually none in vision	unlimited perceptual access, especially in vision; spatial structure useful as an organizational device

Source: From Donald, M. (1991, p. 315).

(b) Cognitive reorganization. External memory has introduced new cognitive skill-clusters that are generally referred to as "literacy" skills, but full symbolic literacy extends well beyond the traditional Western definition of literacy, that is, alphabetic reading competence. The neuropsychology of various acquired dyslexias, dysgraphias, and acalculias has revealed a cluster of functionally dissociable cognitive "modules" in the brain that are necessary to support these skills (see, for instance, Morton 1984; Shallice 1988 [see also multiple book review: *BBS* 14(3) 1991]; Shallice & Warrington 1980).

The localization of these neural modules seems to vary across individuals, as might be expected, since the whole structure must have been imposed by cultural programming rather than by any specific genetic predisposition built into the nervous system. There is a great deal of evidence in single-case neurological histories that these "literacy support networks" are anatomically and functionally distinct from those that support oral-linguistic skills, as well as from those brain regions that support basic perceptual and motor functions (see especially Shallice 1988).

There are at least three dissociable high-level visual interpretative paths involved in symbolic literacy. The most basic is "pictorial," and is needed to interpret pictorial symbols such as pictograms and visual metaphors; even at this level there are numerous interpretative (mostly metaphoric) conventions to master. The second is "ideographic," and is sometimes called the direct visual-semantic path in studies of reading (see Coltheart et al. 1980; Paradis et al. 1985); it maps visual symbols directly onto ideas, as in the case of Chinese ideographic writing, most systems of counting, or many street signs and analog graphic devices like maps, histograms, and charts. The third is "phonetic," and serves to map graphemes onto phonemes, as in alphabetic print. These three paths emerged at different historical phases of visuosymbolic evolution and remain functionally independent of one another; moreover, each path feeds into a distinct visual "lexicon" of thousands of recognizable symbols.

(c) Physical basis. How could the highly complex functional subsystems necessary for reading, writing, and other visuosymbolic processing skills be accommodated by the human brain without genetic change? The answer seems to lie in the increased neocortical plasticity that came with the final expansion of the human brain. This increase in plasticity might be partly a function of greater cortical asymmetry, which allows nonredundancies of function between homologous association regions in the two hemispheres, in effect creating twice as much "extra" neuronal space as a comparable expansion in primary cortical regions, which tend to be more symmetrical in function. Moreover, the immense tertiary neo-neocortical subregions of the human brain have so many competing input pathways that epigenetic factors such as those described by Changeux (1985) and Edelman (1987) could create a very great range of potential functional arrangements. In effect, it is probably because of the plasticity of this arrangement that the human brain can invest so heavily in the decoding baggage needed for using large numbers of novel external memory devices.

In addition, there is evidence that even in adults the cerebral cortex is constantly readjusting and fine-tuning its assignment of processing space, reflecting the constantly changing use patterns imposed by the environment. The somatosensory regions of neocortex may be reorganized by a prolonged increase in stimulation; in fact, the area dedicated to fine touch discrimination expands and contracts in response to imposed load changes (see, for instance, Merzenich et al. 1987). This sensitivity to use pattern may even extend to functions quite different from those that normally occupy a given region, as in the case of the auditory cortex of a congenitally deaf person, which in the absence of auditory stimulation will eventually assume visual functions (Neville 1990). If the relatively hard-wired primary sensory regions are this flexible then tertiary cortical regions ought to show even greater flexibility in their function, given the additional degrees of freedom added by moving two synapses or more from the many sensory, motor, and association regions that impinge upon them.

There appear to be tradeoffs inherent in this flexible arrangement – that is, "invasions" of a given region by an environmentally or culturally driven function will displace other functions that may potentially have depended on that region. This suggests that high levels of literacy skill may entail considerable costs, as indeed has been suggested by the literature comparing the cognitive competences of oral cultures with those of literate ones. Oral memory and visual imagery are often listed among the skills that may have been traded off against literacy (Cole et al. 1971; Richardson 1969).

(d) Changed role of biological memory. One of the most interesting effects of external memory devices is the way they alter human working memory. Working memory is generally conceived as a system centered on consciousness; and although there are many alternatives in the literature, Baddeley's (1986) model was adopted for the purposes of this discussion because it is fairly representative and maps very well onto a neurobiological model.

The tripartite working memory structure proposed by Baddeley includes two slave memory systems, the articulatory loop and the visual-spatial sketchpad, and a central executive. According to this model, when we think, we either imagine, via the sketchpad, or verbalize, via the articulatory loop (the latter may be covert, in the form of "inner speech"). The central executive handles the intermediate-term semantic context – for instance in a conversation it might keep track of what was said, by whom, and in what context. This working memory structure provides the basis for consciousness, although not everything held in working memory is consciously experienced; rather, it is easily available to consciousness.

In preliterate cultures this arrangement, or something close to it, was all that humans had to work with, and its limitations are well documented. A society that relies on this type of memory mechanism would accordingly have to depend upon a variety of social arrangements and mnemonic skills to maintain its accumulating knowledge base: rote verbal recitation, preferably in groups; specialized individuals whose task was to learn and retain knowledge (for instance, shamans and bards); formulaic recital by individuals in an undisturbed, special place; rigidly formal and repetitive group ritual; and various forms of visual imagination as a means of understanding and retaining quite complex memories.

This situation has changed with the increased use of external symbolic storage. The larger architecture within which the individual mind works has changed; in fact, the structure of internal memory is now reflected in the external environment: there is now an external memory field that serves as the real "working memory" for many mental operations, and there is also an external "long-term" store. The external equivalent of the long-term store has very different storage and retrieval properties from those of our internal long-term store. Similarly, the external working memory field has completely different properties from the internal working memory system. Whenever an individual is "plugged in" to some part of the external store, that interaction is mediated by certain items displayed in the external memory field; the latter may consist of a variety of display devices, including print, graphs, monitors, and so on, usually arranged in visual space. The conscious mind is thus juxtaposed between two memory structures, one internal and the other external.

The external display projects directly to the visual regions, which now become the locus of a new kind of internal working memory, one which utilizes the power of the perceptual systems. In effect, because the perceptual systems are displaying *representations* (as opposed to nonsymbolic objects), the user's brain can move through "information space" just as it has always moved through the natural environment, with the difference that processing occurs on two levels, instead of only one. The items displayed in the external memory field are treated first as natural objects and events and second as memory representations that can externally program the user's brain, that is, create specific states of knowledge that were intended by the creator of the particular external device on display.

This second level of analysis, which is the prerequisite for literacy, imposes a great load on visual as well as semantic processing. The process of reading a book, where meanings literally pop out of the script (or the graphs, numbers, ideograms, or other types of symbols it uses) requires a tremendous amount of additional high-level processing. This second level of processing, wholly automated in the expert reader, requires rapid access to thousands of internally stored pictorial, ideographic, graphemic, and visual-lexical codes, along with various specialized grammars, scanning conventions, and a great deal of interpretative knowledge. In effect, this second-level visual system produces knowledge states that are directly driven from the external memory field; it thus becomes the internal display device for a very complex external memory trace. The literate brain thus becomes externally programmable.

But unlike the constantly moving and fading contents of biological working memory, the contents of this externally driven processor can be frozen in time, reviewed, refined, and reformatted. Moreover, all of this can be done intentionally, online, and in real time, in constant interaction with the external display mechanism. In biological working memory, the possibility of this kind of iterative refinement of mental representations is very limited. Neither of Baddeley's (1986) slave systems can support such reflection: the articulatory loop needs constant rehearsal and has a decay time of a few seconds, whereas the visual sketchpad is even more ephemeral, vaguely defined and vulnerable to interference. The central executive is able to hold quite a large amount of information, but in order to consciously modify that information, its contents apparently need to be displayed in one of the slave systems, usually in a covert manner (for instance, inner speech). This imposes a serious limitation on the amount of conscious reflection that can be done on any material that is stored exclusively in biological working memory.

Breaking out of this limited working memory arrangement in itself was a very major change. But it potentiated another important new development: new metalinguistic skills, which expressed themselves in the kinds of symbolic products and cognitive artifacts (Norman 1990) humans could produce and maintain. Producing a single new entry in the external storage system is not a trivial occupation; it never has been, from Ice Age cave paintings to modern science. As artifacts have become much more complex, and the knowledge environment itself has grown, the specific skills needed to become a serious "player" has also taken much more specific preparation, in the form of extended apprenticeships and higher education. There is a trend in the kinds of "metalinguistic" thought skills that have been taught in Western academies over the past few thousand years, moving from an early emphasis on oral and narrative skills, toward visuosymbolic and paradigmatic skills. Denny (1991) has suggested that the major new thought pattern attributable to literacy is a property called "decontextualization," and Olson (1991) has suggested that writing allowed the "objectification" of language, and consequently the development of formal thought skills. These proposals are compatible with my suggestion that literacy allowed the thought process itself to be subjected to iterative refinement through its stable display in the external memory field, and its subsequent incremental refinement, like any other external symbolic product.

The modern brain must accommodate not only these new working memory arrangements and all the coding demands imposed by symbolic literacy, but also a number of metalinguistic skills that simply did not exist a few thousands of years ago. The latter are socially entrenched; for example, an enterprise like modern science is very much a collective endeavor, in which the individual mind is essentially a node in a larger networked structure supported by external memory. Humans have been part of a collective knowledge enterprise ever since mimetic skill permitted us to break with the limitations of episodic cognition, but external memory has amplified the number and variety of representations available in human culture and increased the degree to which our minds share representations and rely on external devices for the process of thought itself. Cognitive studies of the modern workplace (e.g., Hutchins 1990; Olson & Olson 1991; Suchman 1987) testify to the way that electronically distributed knowledge representation, and computer-coordinated planning and problem-solving, are affecting the relative roles of individual minds and external memory devices in this collective enterprise.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged. All page references are to Donald's Origins of the Modern Mind *unless otherwise indicated.*

From cooperative computation to man/machine symbiosis

Michael A. Arbib
Center for Neural Engineering, University of Southern California, Los Angeles, CA 90089–2520
Electronic mail: *arbib@pollux.usc.edu*

In an early target article in *BBS*, Arbib and Caplan (1979) offered a view of neurolinguistics that is in many ways concordant with that offered by Donald in Chapter 3, "Wernicke's Machine." However, where Donald turns to Fodor's (1983) "modules" – which are so large that "language" occupies a single module – we offered a finer-grained analysis that uses the "cooperative computation" of interacting entities called schemas to reveal the salient patterns of interaction embedded in the neural network dynamics of the brain. This intermediate-grain approach to modularity (in the general sense that antedates Fodor) not only encourages computational modeling and neuropsychological analysis of such Fodorian modules as language and vision but it also allows one to embark upon a similarly insightful analysis of the "central processes" that Fodor argues are unanalyzable (Arbib 1987).

Donald, although he uses Fodor's large-scale notion of modules, does take the argument further by analyzing the "central processes" in terms of evolutionary stages. To the basic mechanisms for exploiting episodic memory he adds (1) the "mimetic controller" (Fig. 6.1, p. 190), (2) the "linguistic controller" at the peak of a vertically integrated speech system (Fig. 7.2, p. 260), and (3) the metalinguistic skills which coevolve with the development of increasingly sophisticated "external memory fields." Such "large-module" language leads Donald to characterize the second transition as "one vertically integrated adaptation, ultimately unified under a 'linguistic controller,'" (p. 259) even though elsewhere he reminds us that: "Just as in the case of mimesis, the language adaptation had to involve many different parts of the brain. . . . Once again we are looking at mosaic evolution" (p. 261). (One may note, for example, that a key point in the use of language is to negate a statement and then gather evidence as to which of the alternatives is true. Another is the flexible expression of goals and the ability to analyze various paths to attain them.)

Donald's argument may be strengthened if "schemas" are seen as the units on which evolution may act. Fodor's "modules" then disappear, to be replaced by patterns of schemas which provide a coherent style. As Arbib and Hesse (1986, p. 50) note, "Though processes of schema change may certainly affect only a few schemas at a time, such changes may 'cohere' to yield a mental 'phase transition' into a pattern of holistic organization. [Such transitions may include] stage changes in the sense of Piaget [and] paradigm changes in the sense of Kuhn."

To these examples I would here add "evolutionary transitions in the sense of Merlin Donald." This forces us to spell out more carefully a view of evolution as a form of punctuated equilibrium that goes somewhat as follows: Existing species have reached a local quasi-optimum of fitness in relation to their ecological niche. By quasi-optimum, I mean that the expected effect of a random mutation is a decrease or negligible change in fitness. Thus the species can remain stable for long periods of time until there is either (1) a catastrophic change in the environment so that species fitness is no longer optimal or (2) a very rare mutation does occur that yields a heritable improvement in fitness. The key point is that a successful mutation does not yield a new quasi-optimum. Rather, many different mutations can now effectively yield changes that increase fitness. The biological changes in both bodily and neural structure and function are manifested in new schemas which provide the "mental phenotypes" on which selection acts. We can then seek to see how these changes might be small enough in the genetic metric to be the plausible result of mutations, yet large enough in their functional expression to yield adaptive improvements that can build upon one another to yield a coherent pattern of change. (I would appreciate pointers to scholarly treatments that either support or refute this approach.) Such incremental changes in brain and body are not "unidirectional." Arbib (1989, sect. 7.2) gives a concrete account of the "evolution" of schemas in a computational model for visual motion perception that has strong resonances with a Jacksonian view of brain evolution (Jackson 1874; 1878/1879). This example stresses that "older systems" are not fully encapsulated but can themselves further evolve to take advantage of changes in the "informational environment" afforded by the new brain regions.

Donald's work thus poses two challenges: to understand why the transitions to each of episodic, mimetic, and mythic culture yielded relative stability after a multitude of coadaptive changes (over 50 to 100 generations?) had cohered to yield a new "species style"; and to understand why what Donald calls theoretic culture (marked by man/machine symbiosis; cf. Arbib 1973; 1982) is a stage of explosive cultural change. I do not know the answer, but I think it worth recalling the extension of schema theory that Hesse and I introduced to place "schemas in the head" (cf. stage changes in the sense of Piaget) and "schemas in society" (cf. paradigm changes in the sense of Kuhn) in an integrated perspective.

> An "ideology" is expressed within the whole structure of interactions among the individuals of a society, their institutions, and their artifacts; it can only be vicariously and imperfectly represented within the head of any individual within that society. . . . we use the term "social schema" to denote any such network, whether an ideology, a language, or a religion. . . . Such schemas are not external, like the physical world, but they shape the development of our